READING

Protection of Human Subjects: The Belmont Report and the Common Rule

Before carrying out any research on human subjects, we need to follow certain regulations to ensure that we do not harm our subjects. These regulations are based on a document called the Belmont Report, which was prepared by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979. The Belmont Report provided the foundation for federal regulations pertaining to human subjects research. In 1991, seventeen different federal agencies adopted the Common Rule, a set of federal regulations governing research with human subject.

The Belmont Report

This report outlines three ethical principles for research on human subjects and provides guidelines for assuring that these principles are met:

- **Respect for Persons.** This principle asserts that research subjects have the right to make their own decision about whether they will participate in a research study, or, if they have diminished capacity to make the decision, that a third party make the decision on their behalf. To fulfill this principle, researchers are required to inform their research subjects or their guardians about the purpose and procedure of the study, as well as potential risks and benefits, usually with a written consent form. They need to give subjects the opportunity to ask questions before and during the study and to withdraw at any time during the study. The potential subject or the guardian provides informed consent by signing the consent form. Minor or incapacitated subjects should be given as much information about the study as they can understand and have the opportunity to give assent or decline to participate. It is also important that people not be **coerced** into participating (i.e. suffer consequences if they don't participate) or receive **undue influence** (an excessive reward that might influence someone to participate).
- **Beneficence.** This principle includes minimizing harm to the subjects while maximizing benefits resulting from the study. In planning the research study, the researchers need to consider potential risks to the research subjects, how to minimize the risks, and whether the risks are justified because of the benefits of the study to individuals and society.
- **Justice.** The basis of this principle is that the benefits and burdens of research are fairly distributed. It requires that subjects selected for a research study be from the same population as those who will benefit from the study. In addition, vulnerable populations such as people who are institutionalized, economically disadvantaged, or very sick should not bear an unfair burden of research simply because it is more convenient for the researcher to work with them.

The Common Rule

The Common Rule is a set of federal regulations governing research with human subjects based on the Belmont Report. The Common Rule was adopted by 17 federal agencies in 1991. It establishes the following important regulations:

- Research institutes must assure the federal government that they will enforce protection of human subjects in research done at their institution
- Research institutes are required to have an **Institutional Review Board** (IRB) to review federally funded research studies with human subjects
- Federally funded researchers need to meet requirements for protection of human subjects, including obtaining and documenting informed consent

The role of the IRB is to carefully review all federally funded research studies that use human subjects to make sure that they meet federal regulations. Prior to starting an investigation with human subjects, researchers need to submit a detailed description of their study protocol to the IRB, as well as materials they will use to recruit potential subjects and consent forms they plan to give to their subjects. After reviewing the application, the IRB may approve it, make recommendations for changes, or deny

B2. The *StarNet* Case Control Study Student Sheet [redacted]
(Reading)

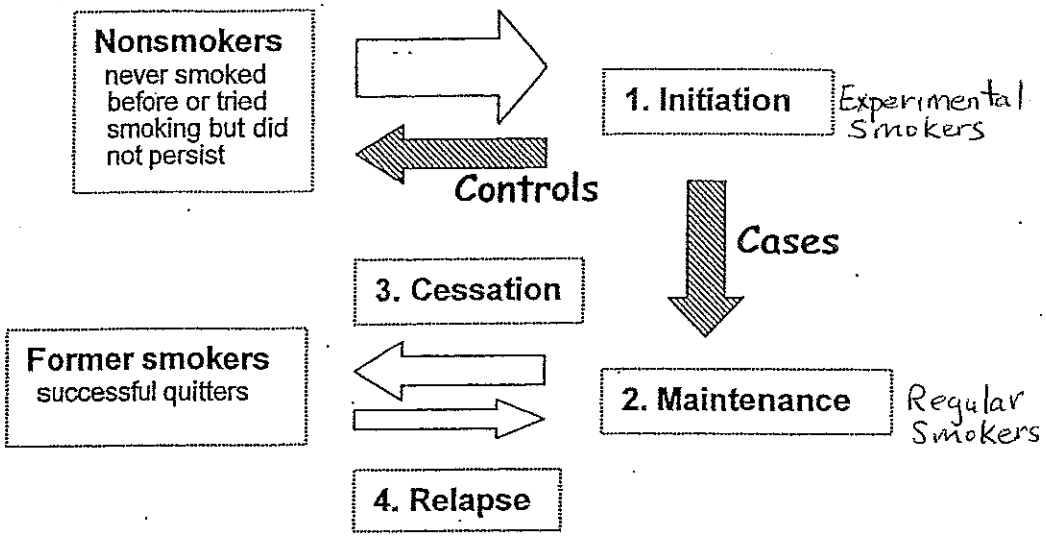
Study Design
The focus of this research study is to investigate the genetic and environmental factors associated with a person moving from smoking initiation to maintenance. Phase 1 of the project involves designing the study. The subjects will include men and women aged 25 to 54 from different ethnic backgrounds (Asian, African American, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, Hispanic or Latino(a), and White, non-Hispanic). Our study will include 400 subjects, with approximately equal number of cases and controls. The case subjects will be smokers who meet several criteria for being persistent tobacco smokers, such as smoking at least 10 cigarettes a day. The controls will be people who've tried smoking a few times but never became persistent smokers.

Why does our study focus on persistent smoking? Many adolescents try smoking—according to the *Monitor the Future* study, 50% of the 12th graders surveyed in 2005 had smoked in their lifetime—but not everyone who tries it becomes a persistent smoker. Why do some people continue to smoke after trying it, while others do not? Answering this question could lead to the design of more effective educational programs or early interventions. As high school students, you are of the age when many people become persistent smokers, so you may have insights into the contributing factors that have not occurred to older scientists.

As you have learned, a case control study is used to figure out what exposures are associated with a particular outcome. In our study, the outcome is transitioning from smoking initiation to maintenance, and the exposures are the genetic and environmental factors that are associated with someone becoming a persistent smoker. We'll take a blood sample from the subjects so we can genotype them for a few candidate genes, like the ones you discussed in Activity A3, *How Genes Influence Smoking Behavior.* We'll use a questionnaire to learn about environmental exposures. As shown in the following *Stages of Smoking* diagram, our case subjects will be persistent smokers, and our controls will be people who have tried smoking but did not continue. Why don't we include all non-smokers in the control group? As shown in the diagram, many people might be considered non-smokers, including those who never tried a cigarette, people who tried but never persisted, and former smokers who have quit. Since our study focuses on the process that leads from initiation to maintenance, our controls should be people who, like the case subjects, have tried smoking, but unlike the cases, did not continue to smoke.

Subjects will be recruited from the Puget Sound area. They will range in age from 25 to 54, because most people initiate smoking before the age of 25, and many people have not quit by the age of 54. Our study will include both men and women from the many ethnic backgrounds represented in the Puget Sound area. Because we will have a diverse study population, it will be critical to ensure that our case and control groups are matched— that they have a similar representation of the different ethnic groups and both men and women within our target age range—so that observed differences between cases and controls are not caused by these demographic differences.

Stages of Smoking



The questionnaire will ask a variety of questions, some of which will have specific purposes in the study. A few demographic questions, such as the subject's age, sex, and ethnicity will be used for matching subjects and controls. Similarly, we will ask a few questions related to the person's smoking habits to sort subjects into cases and controls. The remaining questions will be related to environmental exposures that may be associated with becoming a persistent smoker. Your contribution to this project will be to design questions about environmental exposures. Hopefully you will have ideas about factors that contribute to smoking behavior that scientists have not considered in previous research studies.

The subjects in our study will be competent adults who willingly provide consent after learning the details of the study. Even after giving their consent, they can choose not to answer all the questions on the questionnaire. They will receive a small amount of cash to compensate them for participating, and they will get free parking.

Potential risks in our study include discomfort from giving a blood sample or answering personal questions, and concern about what genetic information will be revealed and its implication. To avoid discomfort from answering the questionnaire, questions will be carefully worded in non-judgmental language, and subjects will have the option of leaving questions unanswered. The blood draw will be done by an experienced and qualified phlebotomist. We will also assure subjects that all data will be anonymous, and no one will be identifiable in the database based on their responses to the questionnaire. Our genetic analysis will focus on common alleles that are not associated with disease conditions. Even if we discover that a particular allele is associated with persistent smoking, no one will be able to identify the subjects who have that allele.

Our study population will be representative of the population of the Puget Sound area. It will consist of competent adults who freely volunteer to participate.

Prior to submitting our application to the IRB, the *StarNet* team will ask scientists to review our study design, consent forms, and recruitment materials, and we will revise them based on their recommendations. After reviewing our application, the IRB may give their approval or make recommendations for modifying it. If changes are required, we will make the necessary revisions and re-submit our application for approval.

Carrying out the Study

Once we have IRB approval for this study, the *StarNet* team will recruit research subjects from around Puget Sound, using newspaper ads and flyers placed at different sites. People who are interested in being part of the study will contact the *StarNet* team by calling a toll free number. A team member will screen each caller to ensure that he or she is suitable for the study. Screening includes finding out whether the caller matches our criteria for being either a case or a control subject. It also involves making sure that we have good representation of men and women of different ethnicities and ages within our target range in both the case and the control group. We want our two groups to be roughly matched in terms of the sex, age, and ethnicities of the subjects.

We will meet potential subjects at a private office. At the beginning of the appointment, the subject will read the consent form and have the opportunity to ask questions about the study. Once his questions have been answered, he may sign the consent form to show his willingness to participate in the study, or, if he does not want to participate in the study, he may leave. Subjects who give their informed consent will answer the questionnaire and have a blood sample drawn.

Phase 2 of the study includes genotyping the subjects' DNA for several candidate genes of interest and entering all the data from the genotyping and the questionnaire into a queriable database. A qualified lab at the UW will prepare genomic DNA from the subjects' blood samples. Then a researcher on the *StarNet* team will amplify small regions of the subjects' DNA using a process called the polymerase chain reaction (PCR). We will send the PCR-amplified DNA out to participating classrooms for genotyping analysis by the students, who will use protocols and materials we provide. The classrooms will submit their results to the *StarNet* team, and we will enter them into the database.

Phase 3 of the research study is analysis of the data, which will be carrying out in high school classrooms. Students will propose hypotheses about factors that might contribute to smoking behavior and test them using data in the database. As students make new discoveries from their analyses of the data, we will post their findings on the project web site so that others can comment on their results or extend them by asking related questions. Hopefully the *StarNet* case control study will shed new light on the question of how genetic and environmental factors affect smoking behavior.

Student Sheet C.1 The Research Questionnaire

In Activity A2, you examined the profiles of several smokers. You probably noticed that many of them shared similar environmental exposures (factors), like peers or relatives who smoked. How do you know which of these exposures influence smoking behavior? Scientists often use research questionnaires to learn more about how environmental exposures are associated with a particular outcome, like becoming a persistent smoker.

The research questionnaire is a tool used by epidemiologists to collect data from their research subjects. Like any research tool, a questionnaire needs to be designed with extreme care so it answers the research question(s) under study. Researchers usually only have one chance to get responses from their subjects, so it's critical that the questionnaire asks for all the information needed to complete the study. Here are some of the things they consider when they write the questionnaire:

1. What research questions do we want to answer with the study?
2. What are our specific research hypotheses?
3. Who are the subjects?
4. What characteristics define the cases? What characteristics define the controls?
5. What kind of questionnaire is it (phone, interview, written, etc.)?

The main research question addressed by the *StarNet* case control study is "What genetic and environmental exposures are associated with people moving from smoking initiation to persistent smoking?" How will we answer that question? First we need to develop reasonable hypotheses about factors that might contribute to becoming a persistent smoker, based on the published literature and our own observations about smoking behavior. Then we need to write questions for the research questionnaire that will provide the data we need to test our hypotheses, especially data on environmental exposures. (Data about genetic exposures will come from our genotyping experiments.)

The *StarNet* questionnaire will include questions from three different categories—smoking behavior, demographic information, and environmental exposures. These questions will be used for four purposes:

- Selecting participants for the study
- Assigning people to the case or control group
- Matching cases and controls
- Determining which environmental exposures are associated with becoming a persistent smoker.

Some of the questionnaire questions will also be asked during a phone interview in which potential subjects are screened to see whether they meet the criteria to be included in the study and assigned to case or control groups. Table C.1 describes each type of question and how it might be used in the study.

Table C.1. Types of questions in the *StarNet* questionnaire

Question category	Types of questions in category	Potential use of questions
Smoking behavior	• Number of cigarettes smoked in lifetime • Number of cigarettes smoked per day • Time of first cigarette each day	• Selecting participants for the study • Assigning to case or control group • Determining whether associated with becoming a persistent smoker
Demographic Information	• Age • Sex • Ethnicity • Level of education • Employment Status	• Selecting participants for the study • Assigning to case or control group • Matching cases and controls • Determining whether associated with becoming a persistent smoker
Environmental Exposures	• Smoking behavior of parents • Smoking behavior of peers	• Determining whether associated with becoming a persistent smoker

Demographic Information
Participants in the *StarNet* study will be asked questions about their age, sex, ethnicity, and education levels. Our subjects are men and women aged 25 to 54 from many ethnic backgrounds. Subjects will be screened by a phone interview before they fill out the questionnaire, but many of the demographic questions used to select subjects during the phone interview will be repeated in the questionnaire so that these data are included in the database. Additional demographic questions that aren't part of the phone interview are also included in the questionnaire.

Smoking Behavior
During the phone interview, subjects will be asked several questions about their smoking behavior that will be used to determine whether to include them in the study and whether to assign them to case or control groups. These questions will be included in the questionnaire, along with additional questions about smoking behavior.

It makes sense to use "ready made" questions to determine how much a person smokes, rather than developing our own questions. Standardized questions have been through many rounds of editing and research to improve them, and using them allows us to compare our findings with existing research more readily.

Perhaps the most widely used test to assess nicotine dependence is the Fagerstrom Test for Nicotine Dependence (FTND) (see Figure C.1). This questionnaire is surprisingly short, consisting of just six questions. Responses to this multiple-choice test are assigned a numerical value, and at the end of the

test, these values are added. Lower scores indicate low dependence on nicotine, and higher scores indicate higher dependence. In our study, cases will have a score of 6 or more on the FTND. Controls will be people who tried smoking but never became persistent smokers.

Figure C.1: Fagerstrom Test for Nicotine Dependence

Is smoking "just a habit" or are you addicted? Take this test and find out your level of dependence on nicotine.

1. How soon after you wake up do you smoke your first cigarette?
   - After 60 minutes (0)
   - 31-60 minutes (1)
   - 6-30 minutes (2)
   - Within 5 minutes (3)
2. Do you find it difficult to refrain from smoking in places where it is forbidden?
   - No (0)
   - Yes (1)
3. Which cigarette would you hate most to give up?
   - The first in the morning (1)
   - Any other (0)
4. How many cigarettes per day do you smoke?
   - 10 or less (0)
   - 11-20 (1)
   - 21-30 (2)
   - 31 or more (3)
5. Do you smoke more frequently during the first hours after awakening than during the rest of the day?
   - No (0)
   - Yes (1)
6. Do you smoke even if you are so ill that you are in bed most of the day?
   - No (0)
   - Yes (1)

Environmental Exposures

Throughout a person's lifetime, they are exposed to many different things in their environment. What environmental exposures do you think are associated with persistent smoking?

Questions:

1. Why is it important that the controls in our study have initiated smoking instead of including all non-smokers as controls?
2. What are two environmental factors that you think are associated with becoming a persistent smoker? Discuss why you think there might be an association.

Name:_____ Date:_____ Period:_____

STUDENT SHEET 5.6 **Sources of Error in Case Control Studies**

What are some possible explanations for a strong association between an exposure and outcome but yet few criteria for causality are met?

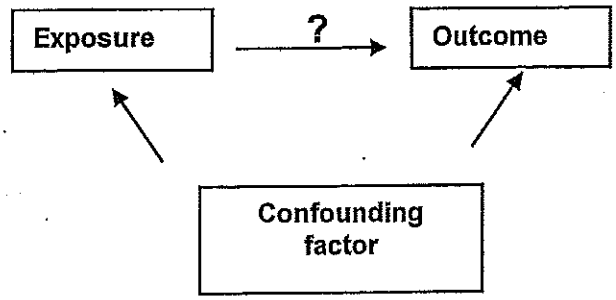
Random error. A detected association could be due to chance. For example, the sample may not be representative of its source population. This type of false association may disappear if the study is repeated with a larger sample size or with a new sample drawn from the same population. When you use the Smoking Behavior database to look for associations, you will be able to check for the possibility of random error with a statistical measure called the **confidence interval**.

Systematic error. An association could be due to a systematic error in the study design. This type of error is called **bias** and does not diminish when the sample size is increased or when the study is repeated with a new sample. When you detect an association using the Smoking Behavior database, you will have to think carefully about whether it could be due to bias. There are several types of bias that are especially important in case control studies:

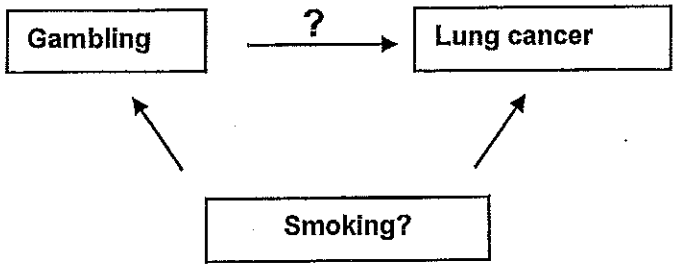
- **Selection bias** occurs when study participants are different than the population from which they are drawn, or when cases and controls come from different populations. Imagine a case control study that looks at whether exposure to car exhaust might cause asthma. The cases are asthma patients in an urban primary care clinic. The controls are patients in a rural dental clinic. Controls may have been exposed to car exhaust less than cases because of their location, leading to an association that may be false. Cases and controls should be drawn from the same area.

- **Information bias** occurs when there is an error in the measurement of the outcome or the exposure for some subjects. One type of information bias is called **recall bias**. This type of bias can occur if cases are better able to recall possible exposures than controls. For example, people with skin cancer may be more likely to recall times when they had extreme sunburns than people without skin cancer. This would result in a stronger association between extreme sunburns and skin cancer than the true strength of association (odds ratio).

- **Confounding** occurs when another factor is associated with both exposure and outcome, giving the appearance that the exposure and outcome are associated (see diagram that follows). Confounding is likely to occur when cases and controls not only differ in the exposure of interest but also in other exposures that may affect the same outcome. In case control studies, careful **matching or restriction** of cases and controls can help reduce possible confounding factors.



Could a confounding factor be associated with both the outcome and the exposure and account for some or all of the observed association between the two?



Gambling has been associated with cancer, but it seems unlikely that gambling would cause cancer (it does not make biological sense). What factors might be associated with both gambling and cancer?

How can we determine whether gambling causes cancer, or if another factor, like smoking, is associated with both gambling and cancer? During data analysis, we can **stratify** (separate) our data based on a suspected confounding factor (e.g. smoking). We would calculate the odds ratio for smokers and non-smokers separately, as shown in the two 2x2 tables below. If the resulting odds ratio is 1 for both groups then there we cannot conclude that gambling is directly related to cancer. .

Smokers

	Cases (have lung cancer)	Controls (no lung cancer)
Exposed (gambling)		
Not exposed (no gambling)		

Nonsmokers

	Cases (have lung cancer)	Controls (no lung cancer)
Exposed (gambling)		
Not exposed (no gambling)		

Confounding can also be minimized by designing the study carefully. We can try to **match** cases and controls for other known risk factors (exposures) for the outcome, or we can **restrict** the study to participants who do not have a suspected confounding factor, for example, by only including non-smokers in the gambling-lung cancer study.

Name:_____ Date:_____ Period:_____

STUDENT SHEET 5.5. **Reading: How do we know if an association is significant or occurred by chance?**

You have learned that an odds ratio (OR) is used to show association between an exposure and an outcome.

> ***Odds ratio**: a measure of the strength of the association between an exposure and an outcome.*

Review the figure below to remind yourself how the odds ratio is calculated.

	Cases	Controls
Exposed	a	b
Not Exposed	c	d

Odds that a case was exposed:

$$\frac{\text{number of cases exposed}}{\text{number of cases not exposed}} = \frac{a}{c}$$

Odds that a control was exposed:

$$\frac{\text{number of controls exposed}}{\text{number of controls not exposed}} = \frac{b}{d}$$

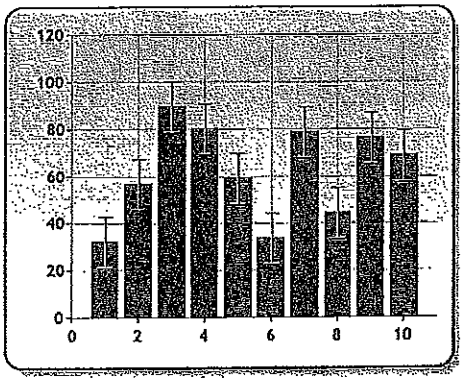
Odds ratio: $\frac{a/c}{b/d}$

"Cases are $\frac{a/c}{b/d}$ more likely to have been exposed than controls"

If we find an association (OR ≠ 1), then we need to determine whether that association is statistically significant or occurred simply by chance. In any case control study, we can only measure a sample of the total cases and controls in the entire population, so when we select subjects for the study, we try to get a good representation of the population. In general, associations that occur by chance are less likely when the sample size is large. Statistics, calculated from samples of the population, allows us to estimate the "true OR" found in the entire population.

STUDENT SHEET 5.5, continued

The 95% Confidence interval = error bar for odds ratios

Traditional error bars	95% Confidence interval
	A range of values that is believed to contain the true population OR with 95% confidence Expressed as: 95% CI = [lower bound, upper bound]
Often determined experimentally	Calculated: $\text{Lower Bound} = OR \times e^{-1.96 \times \left(\sqrt{1/a + 1/b + 1/c + 1/d}\right)}$ $\text{Upper Bound} = OR \times e^{1.96 \times \left(\sqrt{1/a + 1/b + 1/c + 1/d}\right)}$

In our analysis of data from the Smoking Behavior case control study, we will calculate the **95% confidence interval**, which is a range of values for the OR that we are 95% confident contains the true OR of the entire population. Another way to say this is if we repeated the study many times, drawing a different sample from the population each time, 95% of the time the confidence interval would contain the true OR of the population.

> ***Confidence interval:*** *a range of values that is believed to contain the true OR with 95% confidence. It is a measure of how confident we are that the OR in our study reflects the true OR of the population.*

The upper and lower bounds of the confidence interval are calculated using a complex formula involving the values of the OR and a, b, c, and d from the 2 x 2 table. In order for an association to be significant, the 95% confidence interval must not include the value 1. Remember, an OR = 1 means there is no association, so a statistically significant confidence interval should not include this value.

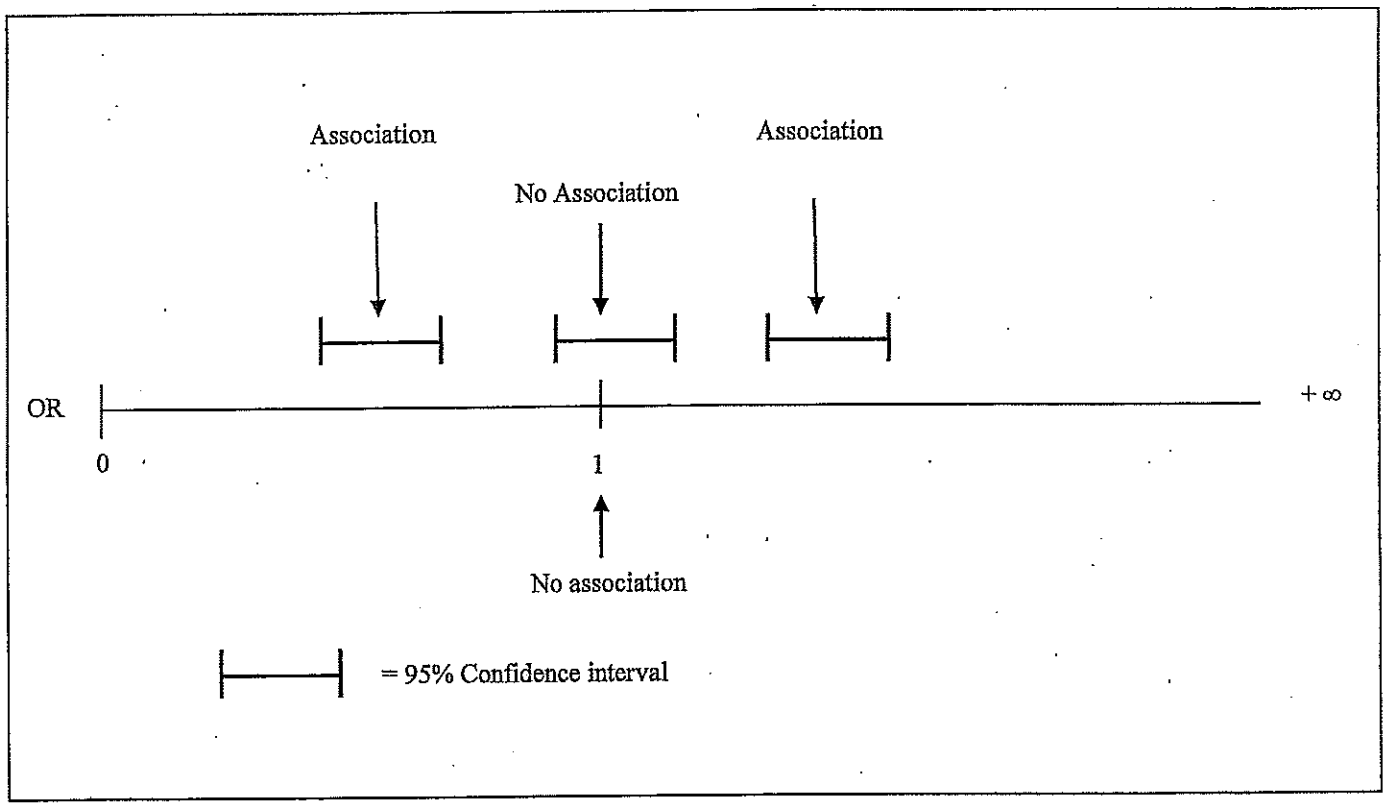
In general, ORs with small confidence intervals are more reliable than ones with large intervals. As the size of the study goes up, the confidence interval size decreases.

STUDENT SHEET 5.5, continued

The formula for the 95% confidence interval is given in the table above. This range of values, along with the OR, will be calculated automatically by the computer when you query the smoking behavior database.

Although you will not need to calculate the confidence interval yourself, you will need to know how the confidence interval is used to determine if the association between an exposure and outcome is significant. If you remember only one thing about the confidence interval, remember that it must not contain an OR=1 if the association is significant.

How to interpret the 95% Confidence interval.



Cheat Sheet: Odds Ratio Output Definitions and Hints

Cases: Regular Smokers

Controls: People who tried or experimented with smoking but never became regular smokers

Exposed: The factor you think might have an influence on someone becoming a regular smoker. For example, believing smoking is not harmful for your health OR having a least 1 parent who smoked.

Not Exposed: The factor you think might protect people from becoming regular smokers. For example, believing smoking is harmful for your health OR not having a parent who smoked.

2 x 2 Table: Is the way people organize data in a case control study. It tell you how many smokers and how many nonsmokers fall into the exposed or not exposed categories. How you define exposed and not exposed depends on how you drag and drop the answers to this question.

Sample Size: The total number of people who responded to this question, both cases and controls.

Odds Ratio: An odds ratio of 1 means there is no association of the exposure between regular smokers and nonsmokers.

95% Confidence Interval: The confidence interval is a tool to help you decide if your result is meaningful to the entire population- people in general. If the confidence interval has the number 1.00 in it, this means that even if your Odds Ratio is bigger or smaller than 1.00, there is not an association between the exposure you identified and regular smokers. However if 1.00 is not inside your Confidence Interval, this means there is an association between the exposure and becoming a regular smoker.

Risk Factor: An exposure (such as believing that smoking is not harmful to one's health) associated with an increased risk of outcome (such as becoming a smoker). Such an exposure is a risk factor when there is an association (OR >1, 95% CI does not contain one) and the causality criteria is fulfilled.

Protective Factor: An exposure (such as bicycle helmets) associated with a reduced risk of outcomes (head injuries). Such an exposure is protective when there is an association (OR <1, 95% CI does not contain one) and the causality criteria is fulfilled.